Queueing Syst (2013) 75:111–113
DOI 10.1007/s11134-013-9376-4

EDITORIAL

Introduction: queueing systems special issue on queueing systems with abandonments

John Hasenbein · David Perry

Published online: 19 September 2013

The investigation of queueing systems with impatient customers has a long history in queueing theory, with the now standard Erlang-A model going back at least to Palm's classic 1957 paper. Despite this long record of work, focus on systems with abandonments has been rekindled due to recent interest in analyzing complex service systems, most notably large call centers. Interestingly, this new surge of work has been motivated both by applied and theoretical advances. In the applied realm, very large call centers are now an important component in many industrial sectors and advanced call management software enables the use of sophisticated algorithms in operating these centers. On the more theoretical side, there have been great advances in approximating large queueing systems with abandonments via many-server asymptotics and related diffusion approximations. Furthermore, there have been numerous advances in the understanding of flexible server models with abandonments, which serve as models for systems with "cross-trained" agents. Techniques from revenue management, game theory, and stochastic optimization have also been applied to the operation of such systems. In addition, queueing systems with impatient customers also have a rich connection with perishable inventory and organ transplantation systems. Thus the requirements of practitioners, and the myriad new tools of theoreticians, have motivated this focused issue on stochastic systems with abandonments.

This special issue actually consists of nine papers, eight of which appear together here. Due to a publishing error, one of the papers, by Baris Ata and Mustafa H.

J. Hasenbein (✉)
Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712-1063, USA
e-mail: jhas@mail.utexas.edu

D. Perry
Department of Statistics, University of Haifa,
199 Aba-Hushi Avenue, Mount Carmel, 3190501 Haifa, Israel
e-mail: dperry@stat.haifa.ac.il

Tongarlak, appeared earlier in a regular issue of the journal (Vol. 74, No. 1, pp. 65–104). The papers in the collection can be roughly divided into three categories. The first paper in the collection surveys the issues that arise from analyzing and modeling real-world systems with impatient customers. The next four papers focus on exact analysis of queueing models with impatience, although the methods vary, ranging from classical Markov chain analysis to stochastic recursion methods. The last four papers emphasize approximate and/or asymptotic analysis of systems with abandonments. In the last 10 years, such analysis has been one cornerstone of resurgent interest in these systems.

The opening paper in our collection, "Data-stories about (Im)Patient Customers in Tele-queues" by Avishai Mandelbaum and Sergey Zeltyn differs from many papers in the queueing literature in that it focuses on empirical data and its relation to appropriate abandonment models. Large data sets collected from a variety of sources give rise to unique issues when customer abandonment is present and these issues are conveyed here through "data-based pictures of impatience." The authors also spend time addressing the important topic of waiting time categorization, for example, expected waiting time versus actual patience time. A bonus feature of this paper is that many of the data sets used are available to researchers in an easily accessible format.

As mentioned above, the next four papers focus on exact analysis of abandonment models. The first of these papers, "On the Time-dependent Moments of Markovian Queues with Reneging" by Brian H. Fralix, analyzes the classic M/M/1+M model (referred to there as the M/M/1−M model). In this paper, Fralix derives new, time-dependent (transient) moments of the queue-length process and also gives new insights into the distribution of the sojourn time in such systems, under the LCFS-PR discipline. The next paper in this group, "Analysis of an M/M/1+G Queue Operated under the FCFS Policy with Exact Admission Control" by Sudipta Das, Lawrence Jenkins, and Debasis Sengupta investigates the interesting policy of *exact admission control*. In this case, the system manager admits only jobs that can be completed within their deadline. For this system, the authors obtain new explicit expressions for the workload distribution, loss ratio, and sojourn time in steady state. Andreas Brandt and Manfred Brandt study an abandonment system with a twist, in their paper "Workload and Busy Period for M/GI/1 with a General Impatience Mechanism." Here the twist is the general impatience mechanism, which allows customers to abandon the system not only while waiting for service, but also while *receiving* service. An advantage of this general model is that it covers the cases of impatience limits of either waiting time or total system time. Their approach uses an ingenious vector process of the virtual waiting time and the age of the current busy period. The last paper in this group "On Queues with Impatience: Stability, and the Optimality of Earliest Deadline First" by Pascal Moyal is unique to this collection in a number of ways. First, the paper uses the framework of stochastic recursions to model the dynamics of the system. This framework allows Moyal to obtain result under very general assumptions on the stochastic primitives (essentially only stationarity and ergodicity are required). Second, the paper investigates the issue of stability, rather than other performance measures. Third, the general framework therein leads to a new rigorous proof of the optimality of the Earliest-Deadline-First policy, with respect the abandonment probability.

The last four papers in our collection develop approximations for systems with abandonment. Such approximations are especially important for practical implementation of queueing results. The first papers operate in the Halfin–Whitt regime, in which the arrival rate and number of servers are scaled up simultaneously. The last two papers operate in the usual heavy-traffic regime, in which only the traffic intensity (but not the number of servers) is scaled. First, William A. Massey and Jamol Pender examine dynamic rate queueing systems in their paper, "Gaussian Skewness Approximation for Dynamic Rate Multi-Server Queues with Abandonment." The feature of this paper that distinguishes it from all other papers in the collection is that the arrival rate is assumed to be non-homogeneous. The other interesting feature of the paper is use of a three-moment Gaussian approximation, involving skewness, which improves upon earlier two-moment approximations. In "Abandonment vs. Blocking in Many-Server Queues: Asymptotic Optimality in the QED Regime," Ananda Weerasinghe and Avishai Mandelbaum investigate an alternative to admission control in systems with abandonment: queue-capacity control. In such systems, there is a tradeoff between blocking customers and having them abandon the system. Operating again in the Halfin–Whitt regime, they are able to provide a formula for the optimal queue capacity that is asymptotically optimal. The third paper in this group "Dynamic Scheduling of a GI/GI/1 + GI Queue with Multiple Customer Classes," by Jeunghyun Kim and Amy R. Ward, is the first in the collection to examine system with multiple customer classes and associated heterogeneous holding costs. Solving a Brownian control problem arising from the heavy-traffic regime leads the authors to propose to a dynamic priority scheduling rule. The interesting point here is that while static scheduling rules work well (sometimes optimally) in multiclass systems without abandonments, they can perform poorly when abandonments are allowed. The last paper in this collection (which appeared in an earlier issue) is "On Scheduling a Multiclass Queue with Abandonments under General Delay Costs" by Baris Ata and Mustafa H. Tongarlak has much in common with the paper by Kim and Ward. Again, they study a multiclass abandonment model with heterogeneous holding costs. The main twist in this paper is that convex and convex-concave holding costs are studied, in addition to the usual case. As before, we note that static index policies are not optimal. Hence, the authors develop effective dynamic indexing policies and in the process create a novel method for solving the Bellman equation associated with the Brownian control problem.

We hope that this collection of papers will spur new research, applied and theoretical, on queueing systems with abandonments and lead us all to new insights into these intriguing queues.